Visualization of Networked Collaboration in Digital Ecosystems through Two-mode Network Patterns

Felix Mödritscher, Wolfgang Taferner

Institute for Information Systems and New Media, Vienna University of Economics and Business

Augasse 2-6, 1090 Vienna, Austria

+43-1-31336-5277

felix.moedritscher@wu.ac.at, wolfgang@taferner.org

Ahmet Soylu, Patrick De Causmaecker

Department of Computer Science, K.U. Leuven
Interdisciplinary Research on Technology, Education and Communication – IBBT, K.U. Leuven

E. Sabbelaan 53, 8500 Kortrijk, Belgium

{Ahmet.Soylu,Patrick.DeCausmaecker}@kuleuven-kortrijk.be

ABSTRACT
Collaboration in Digital Ecosystems can be very complex due to varying types and numbers of actors and artifacts, and the many possible interactions between these entities. Hereby, network visualizations are useful for analyzing networked collaboration and consequently for supporting cognitive processes, like fostering reflection, enabling awareness in students' learning. In this paper, we examine different techniques for visualizing ICT-enabled interactions in Digital Ecosystems. After giving a brief overview of related work, we argue for the application of two-mode networks for visualizing patterns of networked collaboration and discuss different issues by comparing this technique to traditional visualizations.

Categories and Subject Descriptors
I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods: *semantic networks*, E.2 [**Data Structures**]: Data Storage Representations: *linked representations*, H.3.4 [**Systems and Software**]: Information Networks

General Terms
Algorithms, Measurement, Experimentation.

Keywords
Information Visualization, Collaboration, Digital Ecosystems, Two-Mode Networks.

1. INTRODUCTION
Learners and knowledge workers can be seen as actors in a Digital Ecosystem which we understand as "*an open, loosely coupled, domain clustered, demand-driven, self-organising agent environment*" [1]. In such ecosystems learners interact with other actors, communities, artifacts, and tools in order to achieve common goals [2]. Due to large numbers of different entities, interaction flows in Digital Ecosystems can be very complex. Consequently, it is hard to describe and formalize experiences of networked collaboration and learning [1]. In this context, visualizations can be used for analyzing user interactions in Digital Ecosystems, illustrating and explaining characteristics, fostering reflection and awareness in learning, providing pedagogical support etc. [1, 3, 4, 5], whereby normally network visualizations are utilized for these purposes.

However, analysis and visualizations are often restricted to one-mode networks, i.e. the complexity of real-world models of Digital Ecosystems is broken down to networks which contain only one node type. The most popular approach for exploring networked structures in such ecosystems is social network analysis (SNA) which focuses on the relationships (edges) among social entities i.e. humans (nodes) [4, 6]. Other well-known examples are concept maps or tag networks. In any case, the application of one-mode networks always reduces the complexity of networks to one node type, which basically means that information is lost.

In this paper we examine the application of two-mode networks [7] to analyze and visualize patterns in networked collaboration in Digital Ecosystems on the basis of a pattern detection approach. Our assumption is that two-mode networks provide more visual information than one-mode networks and that recurring structural patterns can be identified. The paper is organized as follows. The upcoming section elaborates the theoretical background and gives an overview of related work. Then, we present our pattern-based approach and compare visualizations with other approaches on the basis of various aspects, before the paper is concluded and future work is highlighted.

2. VISUALIZATION OF NETWORKED COLLABORATION IN DIGITAL ECOSYSTEMS

2.1 Visualization Techniques and Network Modes
As mentioned before, Digital Ecosystems and interactions within them can be very complex, e.g. like biological processes in living organisms [1]. In order to analyze such structures, research in the field of Knowledge Discovery is focusing on visualization techniques (linear structures, hierarchies, networks, multi-dimensional spaces, maps, cf. [8]) to be applied for realizing and providing exploration environments, e.g. for discovering relationships, data and document mining, analyzing massive data sets, creating awareness for specific issues, and many more [9].

MEDES'11, November 21-24, 2011, San Francisco, USA.

Due to the emergence of the Web 2.0 and social software in the last few years, the focus has been set to harnessing the collective intelligence in (ICT-based) ecosystems, for instance through data mining and knowledge discovery, and particularly through social network analysis (SNA) [6]. Nowadays, large-scale platforms like Facebook, Google, Bing, the various Apple Stores, Mendeley etc. are analyzing and exploiting user data in every possible way in order to provide value-adding services for their customers. Amongst these services, users also get (visual) feedback on their environments and the ecosystems they are part of. This feedback reaches from simple statistics (e.g. Google Analytics) over exploration facilities (e.g. 'People you may know' browser on Facebook) up to various visualization services (e.g. Microsoft Academic Search or 'My Social Network' by Western Union).

In many cases, these services use network visualization techniques or, at least, are based on networked structures. According to [10], network visualization can "*help users to understand and manage the complexity of large, structured hypermedia collections*". However, it is observable that most of the networks applied in practice are so-called one-mode networks, i.e. networks which consist of nodes of one specific type, e.g. humans (social network), publications (citation networks), tags (tag networks) etc. On the other hand, two-mode networks are based on bipartite graphs with two different node types in which only different nodes are connected with each other [7, 11]. This kind of networks is hardly used in literature and practice.

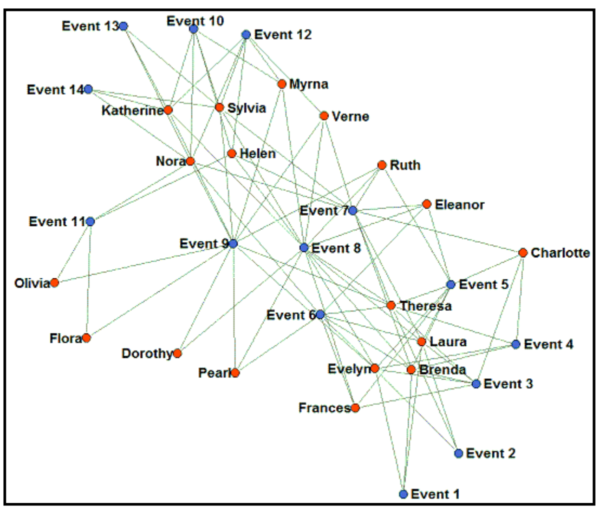

Figure 1. Two-mode network of a clique attending common events (cf. social study in the 'Deep South' project [12]).

A prominent example for analyzing and visualizing two-mode networks is described in [12]. In this study social activities of 18 women (attendance in 14 events) were tracked and visualized through a two-mode network (see Figure 1). Consequently, this two-mode network was transformed into two one-mode networks: (a) women connected through the events they attended, and (b) events connected by the same attendees. The first variant was the basis for the development of affiliation or co-authorship networks. In practice, most one-mode networks are created by reducing more complex k-mode networks to networks with one kind of nodes, which implies that information is lost. Nowadays, research and development has started to consider two-mode networks for analysis purposes again (cf. [13]).

2.2 Pattern Mining in Network Structures

In order to be able to analyze and visualize two-mode networks generated from larger data-sets, we build our approach upon pattern mining. Hereby, a pattern "*describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice*" [14]. In the context of two-mode networks, a pattern can be defined according to relations between nodes as well as thresholds for metrics on nodes and edges (rules).

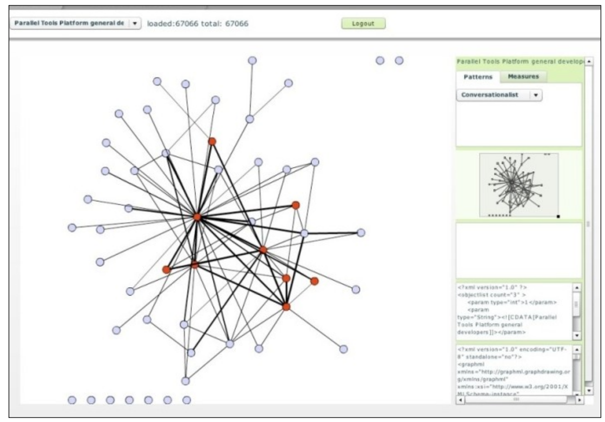

Figure 2. PALADIN visualization of the 'Conversionalist' pattern identified in a forum.

In literature, the use of network structures is very common, for instance for visualizing different kinds of relations (affiliations, co-authorships, interactions, motivational states etc.) between humans (e.g. see [4, 1]). A pattern-based approach is given through PALADIN which stands for 'PAttern LAnguage for DIsturbances in digital Networks' and aims at detecting and visualizing patterns in social media platforms (like forums or mailing lists) [5]. By applying a pattern detection language, it is possible to mine such patterns according to structural and content-based characteristics. Figure 2 gives an example of visualizing a PALADIN pattern. Although the pattern language allows defining very complex patterns the resulting structure and the visualization are always restricted to one-mode networks.

Similarly, other related work is either not capable of generating or processing two-mode networks, or it is simply too costly. For instance, gSpan [15] comprises a very performant technique for graph-based substructure pattern mining. However, gSpan and its algorithmic implementation (e.g. the one for the R framework) are not applicable for bipartite graphs. On the other hand, the 'blockmodeling' approach by [16] is capable of dealing with bipartite (and even k-partite) graphs and allows identifying patterns in such network structures. Yet, the identification of patterns according to this approach requires automated processing of data and therefore is very costly (e.g. the analysis of the study presented in Figure 1, i.e. 18 women and 14 events, caused about 100.000 passes in calculating all possible patterns).

3. COMPARISON OF DIFFERENT NETWORK VISUALIZATIONS

Due to these restrictions of existing approaches and software, we have implemented own algorithms to pre-process data, generate two-mode networks, and visualize patterns of collaboration in Digital Ecosystems. In the following we compare visualization techniques for one and two-mode networks to outline their advantages and possible applications.

3.1 Methodology, Targeted Platforms, and Exploration Environment

For this research we have analyzed different wiki platforms as examples for Digital Ecosystems. Precisely we used the wiki of the SONIVIS project[1] (19 contributors, 342 articles, 600 relations) for defining patterns and two platforms by the Wikimedia Foundation, namely Wikiversity[2] (German) and Wikiquote[3] (German), for validating and refining these patterns on the basis of larger data-sets. After retrieving the raw data from these three Mediawiki instances, the data-sets have been transformed according to a specific data model and have been imported into our exploration environment. As an environment for exploring the networks and developing rules for mining two-mode network patterns we have used the R framework[4], an open source software package for statistical computing and graphics.

3.2 One-mode vs. Two-mode Networks

First of all, we compare visualizations of one-mode and two-mode networks (see also [7]). Figure 3 shows the one-mode network created for the two-mode network shown in Figure 1. Two nodes (women) are directly connected if they attended the same event. It is obvious that the two-mode network includes more information. In this case, it visualizes all events being attended together (cf. Figure 1). The one-mode network only shows that two women met each other at one or more events.

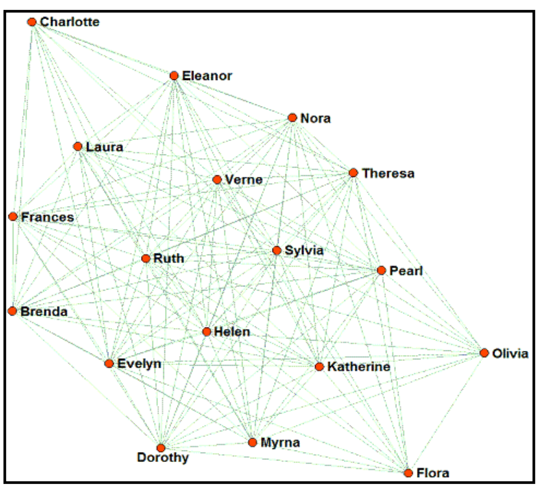

Figure 3. One-mode network generated from the two-mode network displayed in Figure 1.

Basically, selecting the right visualization highly depends on the purpose of its application. If it is necessary to illustrate a certain degree of the complexity of the real world (e.g. the common events being attended by different people) a two-mode or k-mode network would be more useful. On the other hand, visualizing two-mode networks requires more space, which can be problematic if there are too many nodes in the network. As successful Digital Ecosystems tend to have thousands of actors and artifacts, we have decided to go for a pattern detection approach in order to be able to cope with larger networks.

3.3 Pattern Detection in One-mode and Two-mode Networks

We use the PALADIN approach [5] as a starting point for the comparison. Figure 2 shows a network pattern identified within a forum. The actors which are identified as conversionalists – those members in a community who keep the discourse alive – are highlighted with red nodes. Therefore, the pattern ‘conversionalist’ has been specified with a pattern language, taking into account that users of this role start new threads but also post in threads of others, must have a certain number of overall posts. So, basically it is possible to specify such patterns (here: roles of human actors) according to the nodes and metrics, like in our approach. Yet, visualization is always restricted to a one-mode network, the social network.

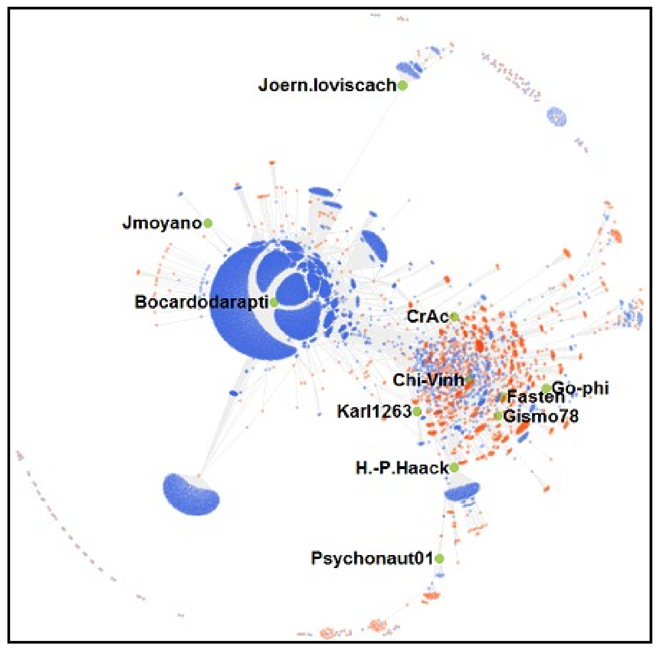

Figure 4. Visualization of all ‘Pioneer’ patterns within the Wikiversity platform.

In contrast to PALADIN, we build upon the contributors and articles of wikis and define the patterns according to these two kinds of nodes and through thresholds for relevant metrics. Figure 4 highlights all ‘Pioneer’ patterns in the Wikiversity data-set (green dots with labels). Thereby, a pioneer is considered to be the designer of a new subject area which is hardly connected to other wiki articles. The pioneer pattern is formalized by rules on the two-mode network, whereby the thresholds for the relevant metrics were determined on the basis of the smaller SONIVIS network and refined with respect to the larger data-sets of Wikiversity and Wikiquote. Our current definition of a pioneer is given by the following rule: A pioneer contributed to at least 0.1%

---

[1] http://www.sonivis.org/wiki/index.php/Main_Page

[2] http://de.wikiversity.org/

[3] http://de.wikiquote.org/

[4] http://cran.r-project.org/

of all wiki articles (to avoid low and non-contributors) and has a 40% to 80% ratio of single-authored to multi-authored articles.

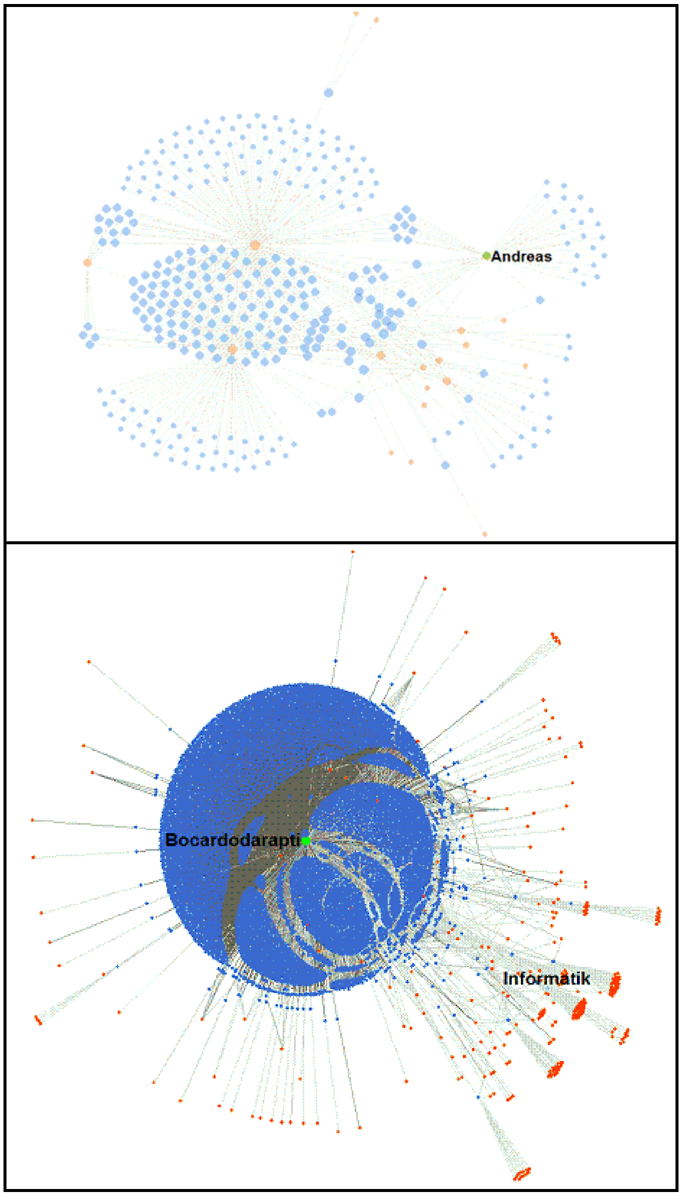

Figure 5. Neighborhood visualization of one ‘Pioneer’ in Sonivis (top) and Wikiversity (bottom).

For developing and analyzing these patterns, however, the visualization of the full network (for our Wikiversity corpora: 1.695 contributors, 9.751 articles and 21.288 relations) is far too detailed and confusing. In order to develop a pattern or analyze parts of the wiki closer, our approach provides the possibility to visualize the neighborhood according to a pattern. Figure 5 shows all collaborators (orange nodes) through which ‘Pioneers’ (green nodes with labels) are connected over their articles (blue nodes), whereby the figure on the top visualizes the collaboration neighborhood of a SONIVIS pioneer (small network) and the one on the bottom presents the neighborhood of a very active user in a large-scale wiki (Wikiversity). Again, it has to be noted that the application of the right visualization should be driven by the purpose of the concrete use case.

3.4 On the Homogeneity of Networked Collaboration in Digital Ecosystems

Finally, we want to briefly address the homogeneity of wikis using our toolkit for analyzing and visualizing networked collaboration. Therefore, we have examined Wikiversity (1.695 contributors, 9.751 articles, 21.288 relations) and Wikiquote (25.911 contributors, 9.120 articles, 136.383 relations), as SONIVIS is too small for such considerations. With the help of two-mode networks and our pattern approach, it can be shown that the collaboration networks vary in terms of homogeneity.

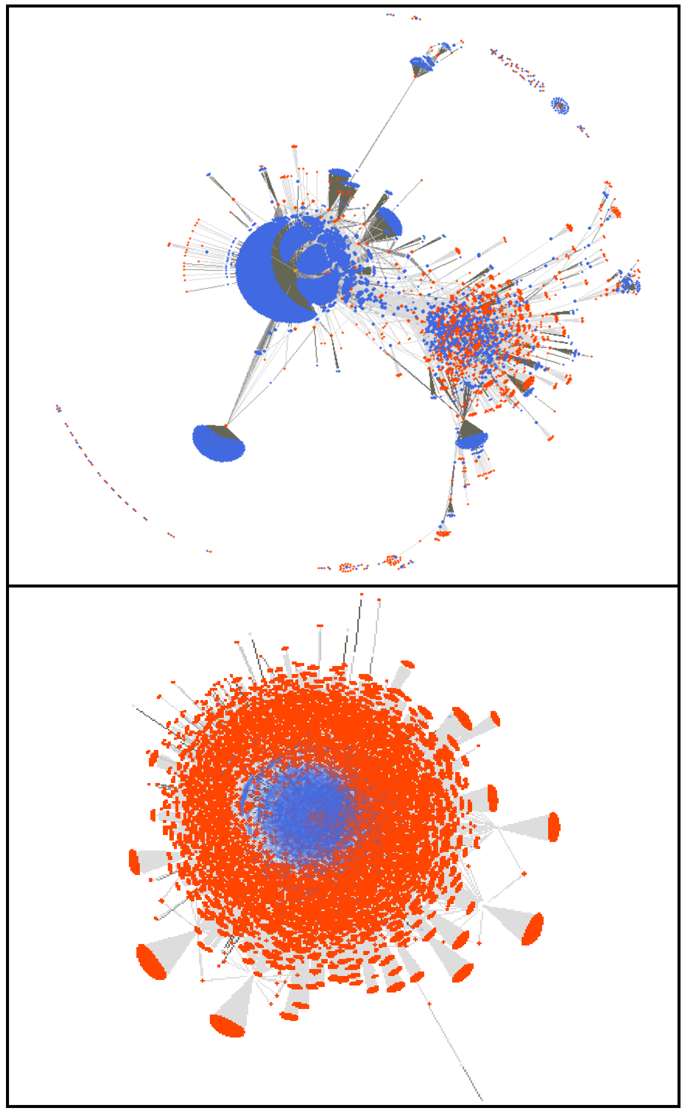

Figure 6. Two-mode network visualization of Wikiversity (top) and Wikiquote (bottom).

In Figure 6 the (full) two-mode networks are displayed next to each other. While the network on the top is very inhomogeneous and consists of various slightly connected sub-networks, the figure on the bottom shows that the relations between contributors and articles are homogeneously distributed over the whole network. Similarly, we have detected no 'Pioneer' pattern for Wikiquote while there are 11 pioneers on the Wikiversity platform (see Figure 4). Vice versa, the pattern 'Networker'– actors connected to a wide range of other contributors – and the pattern 'Community Star' – actors who are well connected through their contributions – occurred more often in the Wikiquote than in the Wikiversity data-set (19:4 and 25:1).

4. CONCLUSIONS AND FUTURE WORK

In this paper we have elaborated benefits of applying two-mode networks for analyzing and visualizing collaboration in Digital Ecosystems. After briefly summarizing the state-of-the-art and related work in this field, we have compared visualizations based on one-mode and two-mode networks generated for a small and two large-scale wiki platforms. Overall, it can be said that two-mode network visualizations include more details on the complexity of real-world structures and processes. On the other hand, two-mode network visualizations might be too overloaded for end-users. A trivial conclusion would be to apply the appropriate network type according to the scenario and the concrete user tasks.

Furthermore, we have shown that a pattern detection approach – like PALADIN for one-mode networks – is realizable for two-mode networks, and that patterns can be identified in two-mode networks. Additionally, a pattern-based approach enables the possibility to zoom into complex network structures, e.g. along the identified patterns. Another important point is that pattern detection can be applied to examine networked collaboration and, for instance, the homogeneity of (wiki-based) communities.

Future work comprises the application of our approach on other wikis – which basically is easy to do for Mediawiki based platforms – or on other technologies, like e.g. personal learning environments (PLEs) and widget technologies [2]. With respect to visualizations, it would be also necessary to realize and evaluate facilities for end-users so that they can explore interactions in Digital Ecosystems. We have not decided the next steps regarding this issue, as the generation of two-mode networks is very costly (processing efforts, time) and the feasibility of providing usable tools and services is unclear.

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) under grant agreement no 231396 (ROLE project).

6. REFERENCES

[1] Giovannella, C., Spadavecchia, C., and Camusi A. 2010. Educational complexity: centrality of design and monitoring of the experience. In *USAB 2010*, LNCS, vol. 6389. Springer, Heidelberg, 353-372. DOI= http://dx.doi.org/10.1007/978-3-642-16607-5_24.

[2] Van Harmelen, M. 2008. Design trajectories: Four experiments in PLE implementation. *Interactive Learning Environments* 16, 1, 35-46. DOI= http://dx.doi.org/10.1080/10494820701772686.

[3] Aviv, R., Erlich, Z., Ravid, G., and Geva, A. 2003. Network analysis of knowledge construction in asynchronous learning networks. *J. of Asynchronous Learning Networks* 7, 3, 1-23.

[4] Korfiatis, N., and Naeve, A. 2005. Evaluating Wiki contributions using social networks: A case study on Wikipedia. In *Proceedings of the first on-Line Conference on Metadata and Semantics Research*.

[5] Klamma, R., Spaniol, M., and Denev, D. 2006. PALADIN: A Pattern Based Approach to Knowledge Discovery in Digital Social Networks. In *Proceedings of the International Conference on Knowledge Management* (Graz, Austria, September 6-8, 2006). Springer, Heidelberg, 457-464.

[6] Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and application*. University Press, Cambridge.

[7] Latapy, M., Magnien, C., and Del Vecchio, N. 2008. Basic Notions for the Analysis of Large Two-mode Networks. *Social Networks* 30, 1, 31-48. DOI= http://dx.doi.org/10.1016/j.socnet.2007.04.006.

[8] Andrews, K. 2002. *Visualising Information Structures: Aspects of Information Visualisation*. Professorial Thesis. Graz University of Technology, Graz.

[9] Fayyad, U., Grinstein, G.G., and Wierse, A. 2002. *Information visualization in data mining and knowledge discovery*. Academic Press, San Diego.

[10] Andrews, K. 1998. Visualizing Rich, Structured Hypermedia *IEEE Computer Graphics and Applications* 18, 4, 40-42. DOI= http://dx.doi.org/10.1109/38.689661.

[11] Saaty, T.L., and Kainen, P.C. 1986. *The Four-Color Problem: Assaults and Conquest*. Dover Publ., New York.

[12] Davis, A., Gardner, B., and Gardner, M. 1941. *Deep South: A Social Anthropological Study of Caste and Class*. University of Chicago Press, Chicago.

[13] Borgatti, S.P. forthcoming. 2-Mode Concepts in Social Network Analysis. In *Encyclopedia of Complexity and System Science*, R.A. Meyers, Ed. Springer, Heidelberg.

[14] Alexander, C., Ishikawa, S., and Silverstein, M. 1977. *A pattern language: Towns, building, construction*. Oxford University Press, Oxford.

[15] Yan, X., and Han, J. 2002. gSpan: Graph-Based Substructure Pattern Mining. In *Proceedings of the IEEE International Conference for Data Mining* (Maebashi City, Japan, December 9-12, 2002). IEEE Computer Society, Washington, DC, 721-724. DOI= http://dx.doi.org/10.1109/ICDM.2002.1184038.

[16] Doreian, P., Batagelj, V., and Ferligoj, A. 2004. Generalized blockmodeling of two-mode network data. *Social Networks* 26, 1, 29-53. DOI= http://dx.doi.org/10.1016/j.socnet.2004.01.002.